

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5

(i) APPLICANT:

- (A) NAME: Akzo nobel n.v.
- (B) STREET: Velperweg 76
- (C) CITY: Arnhem
- (E) COUNTRY: The Netherlands
- (F) POSTAL CODE (ZIP): 6824 BM
- (G) TELEPHONE: 0412-666379
- (H) TELEFAX: 0412-650592
- (I) TELEX: 37503 akpha nl

10

15

(ii) TITLE OF INVENTION: Novel estrogen receptor

(iii) NUMBER OF SEQUENCES: 28

20

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

25

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 1434 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATGAATTACA GCATTCCCAG CAATGTCAC T AACTTGGAAG GTGGGCCTGG TCGGCAGACC 60

ACAAGCCCAA ATGTGTTGTG GCCAACACCT GGGCACCTTT CTCCTTTAGT GGTCCATCGC 120

5 CAGTTATCAC ATCTGTATGC GGAACCTCAA AAGAGTCCCT GGTGTGAAGC AAGATCGCTA 180

GAACACACCT TACCTGTAAA CAGAGAGACA CTGAAAAGGA AGGTTAGTGG GAACCGTTGC 240

10 GCCAGCCCTG TTACTGGTCC AGGTTCAAAG AGGGATGCTC ACTTCTGCGC TGTCTGCAGC 300

GATTACGCAT CGGGATATCA CTATGGAGTC TGGTCGTGTG AAGGATGTAA GGCCTTTTTT 360

AAAAGAAGCA TTCAAGGACA TAATGATTAT ATTTGTCCAG CTACAAATCA GTGTACAATC 420

15 GATAAAAACC GGC GCAAGAG CTGCCAGGCC TGCCGACTTC GGAAGTGTTA CGAAGTGGGA 480

ATGGTGAAGT GTGGCTCCCG GAGAGAGAGA TGTGGGTACC GCCTTGTGCG GAGACAGAGA 540

20 AGTGCCGACG AGCAGCTGCA CTGTGCCGGC AAGGCCAAGA GAAGTGGCGG CCACGCGCCC 600

CGAGTGCGGG AGCTGCTGCT GGACGCCCTG AGCCCCGAGC AGCTAGTGCT CACCCTCCTG 660

GAGGCTGAGC CGCCCCATGT GCTGATCAGC CGCCCCAGTG CGCCCTTCAC CGAGGCCTCC 720

25 ATGATGATGT CCCTGACCAA GTTGGCCGAC AAGGAGTTGG TACACATGAT CAGCTGGGCC 780

AAGAAGATTC CCGGCTTTGT GGAGCTCAGC CTGTTCGACC AAGTGCGGCT CTTGGAGAGC 840

30 TGTGATGATG AGGTGTTAAT GATGGGGCTG ATGTGGCGCT CAATTGACCA CCCC GCAAG 900

CTCATCTTTG CTCCAGATCT TGTTCCTGGAC AGGGATGAGG GGAAATGCGT AGAAGGAATT 960

CTGGAAATCT TTGACATGCT CCTGGCAACT ACTTCAAGGT TTCGAGAGTT AAAACTCCAA 1020

35 CACAAAGAAT ATCTCTGTGT CAAGGCCATG ATCCTGCTCA ATTCCAGTAT GTACCCTCTG 1080

GTCACAGCGA CCCAGGATGC TGACAGCAGC CGGAAGCTGG CTCACTTGCT GAACGCCGTG 1140

ACCGATGCTT TGGTTTGGGT GATTGCCAAG AGCGGCATCT CCTCCCAGCA GCAATCCATG 1200
CGCCTGGCTA ACCTCCTGAT GCTCCTGTCC CACGTCAGGC ATGCGAGTAA CAAGGGCATG 1260
5 GAACATCTGC TCAACATGAA GTGCAAAAAT GTGGTCCCAG TGTATGACCT GCTGCTGGAG 1320
ATGCTGAATG CCCACGTGCT TCGCGGGTGC AAGTCCTCCA TCACGGGGTC CGAGTGCAGC 1380
CCGGCAGAGG ACAGTAAAAG CAAAGAGGGC TCCCAGAACC CACAGTCTCA GTGA 1434

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1251 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

25 ATGAATTACA GCATTCCCAG CAATGTCACT AACTTGGAAG GTGGGCCTGG TCGGCAGACC 60
ACAAGCCCAA ATGTGTTGTG GCCAACACCT GGGCACCTTT CTCCTTTAGT GGTCCATCGC 120
CAGTTATCAC ATCTGTATGC GGAACCTCAA AAGAGTCCCT GGTGTGAAGC AAGATCGCTA 180
30 GAACACACCT TACCTGTAAA CAGAGAGACA CTGAAAAGGA AGGTTAGTGG GAACCGTTGC 240
GCCAGCCCTG TTACTGGTCC AGGTTCAAAG AGGGATGCTC ACTTCTGCGC TGTCTGCAGC 300
35 GATTACGCAT CGGGATATCA CTATGGAGTC TGGTCGTGTG AAGGATGTAA GGCCTTTTTT 360
AAAAGAAGCA TTCAAGGACA TAATGATTAT ATTTGTCCAG CTACAAATCA GTGTACAATC 420
GATAAAAACC GGC GCAAGAG CTGCCAGGCC TGCCGACTTC GGAAGTGTTA CGAAGTGGGA 480

ATGGTGAAGT GTGGCTCCCG GAGAGAGAGA TGTGGGTACC GCCTTGTGCG GAGACAGAGA 540
AGTGCCGACG AGCAGCTGCA CTGTGCCGGC AAGGCCAAGA GAAGTGGCGG CCACGCGCCC 600
5 CGAGTGC GGG AGCTGCTGCT GGACGCCCTG AGCCCCGAGC AGCTAGTGCT CACCCTCCTG 660
GAGGCTGAGC CGCCCCATGT GCTGATCAGC CGCCCCAGTG CGCCCTTCAC CGAGGCCTCC 720
10 ATGATGATGT CCCTGACCAA GTTGCCGAC AAGGAGTTGG TACACATGAT CAGCTGGGCC 780
AAGAAGATTC CCGGCTTTGT GGAGCTCAGC CTGTTCGACC AAGTGGCGCT CTTGGAGAGC 840
TGTTGGATGG AGGTGTTAAT GATGGGGCTG ATGTGGCGCT CAATTGACCA CCCCGGCAAG 900
15 CTCATCTTTG CTCCAGATCT TGTCTGGAC AGGGATGAGG GGAAATGCGT AGAAGGAATT 960
CTGGAAATCT TTGACATGCT CCTGGCAACT ACTTCAAGGT TTCGAGAGTT AAAACTCCAA 1020
20 CACAAAGAAT ATCTCTGTGT CAAGGCCATG ATCCTGCTCA ATTCCAGTAT GTACCCTCTG 1080
GTCACAGCGA CCCAGGATGC TGACAGCAGC CGGAAGCTGG CTCACTTGCT GAACGCCGTG 1140
ACCGATGCTT TGGTTTGGGT GATTGCCAAG AGCGGCATCT CCTCCCAGCA GCAATCCATG 1200
25 CGCCTGGCTA ACCTCCTGAT GTCCTGTCC CACGTCAGGC ATGCGAGGTG A 1251

(2) INFORMATION FOR SEQ ID NO: 3:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Cys Ala Val Cys Ser Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp
1 5 10 15

Ser Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His
20 25 30

Asn Asp Tyr Ile Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn
35 40 45

Arg Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val
50 55 60

Gly Met
65

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Leu Val Leu Thr Leu Leu Glu Ala Glu Pro Pro His Val Leu Ile Ser
1 5 10 15

Arg Pro Ser Ala Pro Phe Thr Glu Ala Ser Met Met Met Ser Leu Thr
20 25 30

Lys Leu Ala Asp Lys Glu Leu Val His Met Ile Ser Trp Ala Lys Lys
35 40 45

	Ile	Pro	Gly	Phe	Val	Glu	Leu	Ser	Leu	Phe	Asp	Gln	Val	Arg	Leu	Leu	
	50					55						60					
5	Glu	Ser	Cys	Trp	Met	Glu	Val	Leu	Met	Met	Gly	Leu	Met	Trp	Arg	Ser	
	65					70					75					80	
	Ile	Asp	His	Pro	Gly	Lys	Leu	Ile	Phe	Ala	Pro	Asp	Leu	Val	Leu	Asp	
						85					90					95	
10	Arg	Asp	Glu	Gly	Lys	Cys	Val	Glu	Gly	Ile	Leu	Glu	Ile	Phe	Asp	Met	
						100					105					110	
	Leu	Leu	Ala	Thr	Thr	Ser	Arg	Phe	Arg	Glu	Leu	Lys	Leu	Gln	His	Lys	
15						115					120					125	
	Glu	Tyr	Leu	Cys	Val	Lys	Ala	Met	Ile	Leu	Leu	Asn	Ser	Ser	Met	Tyr	
	130					135						140					
20	Pro	Leu	Val	Thr	Ala	Thr	Gln	Asp	Ala	Asp	Ser	Ser	Arg	Lys	Leu	Ala	
	145					150					155					160	
	His	Leu	Leu	Asn	Ala	Val	Thr	Asp	Ala	Leu	Val	Trp	Val	Ile	Ala	Lys	
						165					170					175	
25	Ser	Gly	Ile	Ser	Ser	Gln	Gln	Gln	Ser	Met	Arg	Leu	Ala	Asn	Leu	Leu	
						180					185					190	
	Met	Leu	Leu	Ser	His	Val	Arg	His	Ala	Ser	Asn	Lys	Gly	Met	Glu	His	
30						195					200					205	
	Leu	Leu	Asn	Met	Lys	Cys	Lys	Asn	Val	Val	Pro	Val	Tyr	Asp	Leu	Leu	
	210						215					220					
35	Leu	Glu	Met	Leu	Asn	Ala	His	Val	Leu								
	225						230										

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 477 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Asn Tyr Ser Ile Pro Ser Asn Val Thr Asn Leu Glu Gly Gly Pro
1 5 10 15

Gly Arg Gln Thr Thr Ser Pro Asn Val Leu Trp Pro Thr Pro Gly His
20 25 30

Leu Ser Pro Leu Val Val His Arg Gln Leu Ser His Leu Tyr Ala Glu
35 40 45

Pro Gln Lys Ser Pro Trp Cys Glu Ala Arg Ser Leu Glu His Thr Leu
50 55 60

Pro Val Asn Arg Glu Thr Leu Lys Arg Lys Val Ser Gly Asn Arg Cys
65 70 75 80

Ala Ser Pro Val Thr Gly Pro Gly Ser Lys Arg Asp Ala His Phe Cys
85 90 95

Ala Val Cys Ser Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser
100 105 110

Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn
115 120 125

Asp Tyr Ile Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg
130 135 140

Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly
145 150 155 160

5 Met Val Lys Cys Gly Ser Arg Arg Glu Arg Cys Gly Tyr Arg Leu Val
165 170 175

Arg Arg Gln Arg Ser Ala Asp Glu Gln Leu His Cys Ala Gly Lys Ala
180 185 190

10 Lys Arg Ser Gly Gly His Ala Pro Arg Val Arg Glu Leu Leu Leu Asp
195 200 205

Ala Leu Ser Pro Glu Gln Leu Val Leu Thr Leu Leu Glu Ala Glu Pro
210 215 220

15 Pro His Val Leu Ile Ser Arg Pro Ser Ala Pro Phe Thr Glu Ala Ser
225 230 235 240

Met Met Met Ser Leu Thr Lys Leu Ala Asp Lys Glu Leu Val His Met
20 245 250 255

Ile Ser Trp Ala Lys Lys Ile Pro Gly Phe Val Glu Leu Ser Leu Phe
260 265 270

25 Asp Gln Val Arg Leu Leu Glu Ser Cys Trp Met Glu Val Leu Met Met
275 280 285

Gly Leu Met Trp Arg Ser Ile Asp His Pro Gly Lys Leu Ile Phe Ala
290 295 300

30 Pro Asp Leu Val Leu Asp Arg Asp Glu Gly Lys Cys Val Glu Gly Ile
305 310 315 320

Leu Glu Ile Phe Asp Met Leu Leu Ala Thr Thr Ser Arg Phe Arg Glu
35 325 330 335

Leu Lys Leu Gln His Lys Glu Tyr Leu Cys Val Lys Ala Met Ile Leu
340 345 350

Leu Asn Ser Ser Met Tyr Pro Leu Val Thr Ala Thr Gln Asp Ala Asp
355 360 365

5 Ser Ser Arg Lys Leu Ala His Leu Leu Asn Ala Val Thr Asp Ala Leu
370 375 380

Val Trp Val Ile Ala Lys Ser Gly Ile Ser Ser Gln Gln Gln Ser Met
385 390 395 400

10 Arg Leu Ala Asn Leu Leu Met Leu Leu Ser His Val Arg His Ala Ser
405 410 415

Asn Lys Gly Met Glu His Leu Leu Asn Met Lys Cys Lys Asn Val Val
420 425 430

15 Pro Val Tyr Asp Leu Leu Leu Glu Met Leu Asn Ala His Val Leu Arg
435 440 445

20 Gly Cys Lys Ser Ser Ile Thr Gly Ser Glu Cys Ser Pro Ala Glu Asp
450 455 460

Ser Lys Ser Lys Glu Gly Ser Gln Asn Pro Gln Ser Gln
465 470 475

25 (2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 416 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Asn Tyr Ser Ile Pro Ser Asn Val Thr Asn Leu Glu Gly Gly Pro

1	5	10	15
Gly Arg Gln Thr Thr Ser Pro Asn Val Leu Trp Pro Thr Pro Gly His			
	20	25	30
5	Leu Ser Pro Leu Val Val His Arg Gln Leu Ser His Leu Tyr Ala Glu		
	35	40	45
10	Pro Gln Lys Ser Pro Trp Cys Glu Ala Arg Ser Leu Glu His Thr Leu		
	50	55	60
	Pro Val Asn Arg Glu Thr Leu Lys Arg Lys Val Ser Gly Asn Arg Cys		
	65	70	75 80
15	Ala Ser Pro Val Thr Gly Pro Gly Ser Lys Arg Asp Ala His Phe Cys		
	85	90	95
	Ala Val Cys Ser Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser		
	100	105	110
20	Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn		
	115	120	125
	Asp Tyr Ile Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg		
25	130	135	140
	Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly		
	145	150	155 160
30	Met Val Lys Cys Gly Ser Arg Arg Glu Arg Cys Gly Tyr Arg Leu Val		
	165	170	175
	Arg Arg Gln Arg Ser Ala Asp Glu Gln Leu His Cys Ala Gly Lys Ala		
	180	185	190
35	Lys Arg Ser Gly Gly His Ala Pro Arg Val Arg Glu Leu Leu Leu Asp		
	195	200	205
	Ala Leu Ser Pro Glu Gln Leu Val Leu Thr Leu Leu Glu Ala Glu Pro		

	210	215	220
	Pro His Val Leu Ile Ser Arg Pro Ser Ala Pro Phe Thr Glu Ala Ser		
	225	230	235 240
5	Met Met Met Ser Leu Thr Lys Leu Ala Asp Lys Glu Leu Val His Met		
	245	250	255
	Ile Ser Trp Ala Lys Lys Ile Pro Gly Phe Val Glu Leu Ser Leu Phe		
10	260	265	270
	Asp Gln Val Arg Leu Leu Glu Ser Cys Trp Met Glu Val Leu Met Met		
	275	280	285
15	Gly Leu Met Trp Arg Ser Ile Asp His Pro Gly Lys Leu Ile Phe Ala		
	290	295	300
	Pro Asp Leu Val Leu Asp Arg Asp Glu Gly Lys Cys Val Glu Gly Ile		
	305	310	315 320
20	Leu Glu Ile Phe Asp Met Leu Leu Ala Thr Thr Ser Arg Phe Arg Glu		
	325	330	335
	Leu Lys Leu Gln His Lys Glu Tyr Leu Cys Val Lys Ala Met Ile Leu		
25	340	345	350
	Leu Asn Ser Ser Met Tyr Pro Leu Val Thr Ala Thr Gln Asp Ala Asp		
	355	360	365
30	Ser Ser Arg Lys Leu Ala His Leu Leu Asn Ala Val Thr Asp Ala Leu		
	370	375	380
	Val Trp Val Ile Ala Lys Ser Gly Ile Ser Ser Gln Gln Gln Ser Met		
	385	390	395 400
35	Arg Leu Ala Asn Leu Leu Met Leu Leu Ser His Val Arg His Ala Arg		
	405	410	415

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGIGAYGARG CWTGIGGITG YCAYTAYGG

29

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

AAGCCTGGSA YICKYTTIGC CCAIYTIAT

29

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TGTTACGAAG TGGGAATGGT GA

22

10

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

15

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TTGACACCAG ACCAACTGGT AATG

24

25

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

30

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGTGGCGACG ACTCCTGGAG CCCG

24

(2) INFORMATION FOR SEQ ID NO: 12:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTACACTGAT TTGTAGCTGG AC

22

(2) INFORMATION FOR SEQ ID NO: 13:

20

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCATGATGAT GTCCCTGACC

20

35

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

10 TCGCATGCCT GACGTGGGAC 20

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

25

GGCSTCCAGC ATCTCCAGSA RCAG 24

(2) INFORMATION FOR SEQ ID NO: 16:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGAAGCTGGC TCACTTGCTG

20

5 (2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TCTTGTTCTG GACAGGGATG

20

20

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

25 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

35 GCATGGAACA TCTGCTCAAC

20

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGCAAGTTCA GCCTGTTAAG T

21

(2) INFORMATION FOR SEQ ID NO: 20:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1257 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ATGAATTACA GCATTCCCAG CAATGTCAC T AACTTGAAG GTGGGCCTGG TCGGCAGACC 60

30

ACAAGCCCAA ATGTGTTGTG GCCAACACCT GGGCACCTTT CTCCTTTAGT GGTCCATCGC 120

CAGTTATCAC ATCTGTATGC GGAACCTCAA AAGAGTCCCT GGTGTGAAGC AAGATCGCTA 180

GAACACACCT TACCTGTAAA CAGAGAGACA CTGAAAAGGA AGGTTAGTGG GAACCGTTGC 240

35

GCCAGCCCTG TTAGTGGTCC AGGTTCAAAG AGGGATGCTC ACTTCTGCGC TGTCTGCAGC 300

GATTACGCAT CGGGATATCA CTATGGAGTC TGGTCGTGTG AAGGATGTAA GGCCTTTTTT 360

AAAAGAAGCA TTCAAGGACA TAATGATTAT ATTTGTCCAG CTACAAATCA GTGTACAATC 420

GATAAAAACC GGC GCAAGAG CTGCCAGGCC TGCCGACTTC GGAAGTGTTA CGAAGTGGGA 480

5 ATGGTGAAGT GTGGCTCCCG GAGAGAGAGA TGTGGGTACC GCCTTGTGCG GAGACAGAGA 540

AGTGCCGACG AGCAGCTGCA CTGTGCCGGC AAGGCCAAGA GAAGTGGCGG CCACGCGCCC 600

CGAGTGCGGG AGCTGCTGCT GGACGCCCTG AGCCCCGAGC AGCTAGTGCT CACCCTCCTG 660

10 GAGGCTGAGC CGCCCCATGT GCTGATCAGC CGCCCCAGTG CGCCCTTCAC CGAGGCCTCC 720

ATGATGATGT CCCTGACCAA GTTGGCCGAC AAGGAGTTGG TACACATGAT CAGCTGGGCC 780

15 AAGAAGATTC CCGGCTTTGT GGAGCTCAGC CTGTTGACC AAGTGGCGCT CTTGGAGAGC 840

TGTTGGATGG AGGTGTTAAT GATGGGGCTG ATGTGGCGCT CAATTGACCA CCCC GGCAAG 900

CTCATCTTTG CTCCAGATCT TGTCTGGAC AGGGATGAGG GGAAATGCGT AGAAGGAATT 960

20 CTGGAAATCT TTGACATGCT CCTGGCAACT ACTTCAAGGT TTCGAGAGTT AAAACTCCAA 1020

CACAAAGAAT ATCTCTGTGT CAAGGCCATG ATCCTGCTCA ATTCCAGTAT GTACCCTCTG 1080

25 GTCACAGCGA CCCAGGATGC TGACAGCAGC CGGAAGCTGG CTCACTTGCT GAACGCCGTG 1140

ACCGATGCTT TGGTTTGGGT GATTGCCAAG AGCGGCATCT CCTCCAGCA GCAATCCATG 1200

CGCCTGGCTA ACCTCCTGAT GCTCCTGTCC CACGTCAGGC ATGCGAGGTC TGCCTGA 1257

30

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

5
Met Asn Tyr Ser Ile Pro Ser Asn Val Thr Asn Leu Glu Gly Gly Pro
1 5 10 15
10
Gly Arg Gln Thr Thr Ser Pro Asn Val Leu Trp Pro Thr Pro Gly His
20 25 30
Leu Ser Pro Leu Val Val His Arg Gln Leu Ser His Leu Tyr Ala Glu
35 40 45
15
Pro Gln Lys Ser Pro Trp Cys Glu Ala Arg Ser Leu Glu His Thr Leu
50 55 60
Pro Val Asn Arg Glu Thr Leu Lys Arg Lys Val Ser Gly Asn Arg Cys
65 70 75 80
20
Ala Ser Pro Val Thr Gly Pro Gly Ser Lys Arg Asp Ala His Phe Cys
85 90 95
Ala Val Cys Ser Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser
100 105 110
25
Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn
115 120 125
30
Asp Tyr Ile Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg
130 135 140
Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly
145 150 155 160
35
Met Val Lys Cys Gly Ser Arg Arg Glu Arg Cys Gly Tyr Arg Leu Val
165 170 175
Arg Arg Gln Arg Ser Ala Asp Glu Gln Leu His Cys Ala Gly Lys Ala

180

185

190

Lys Arg Ser Gly Gly His Ala Pro Arg Val Arg Glu Leu Leu Leu Asp
195 200 205

5

Ala Leu Ser Pro Glu Gln Leu Val Leu Thr Leu Leu Glu Ala Glu Pro
210 215 220

10

Pro His Val Leu Ile Ser Arg Pro Ser Ala Pro Phe Thr Glu Ala Ser
225 230 235 240

Met Met Met Ser Leu Thr Lys Leu Ala Asp Lys Glu Leu Val His Met
245 250 255

15

Ile Ser Trp Ala Lys Lys Ile Pro Gly Phe Val Glu Leu Ser Leu Phe
260 265 270

Asp Gln Val Arg Leu Leu Glu Ser Cys Trp Met Glu Val Leu Met Met
275 280 285

20

Gly Leu Met Trp Arg Ser Ile Asp His Pro Gly Lys Leu Ile Phe Ala
290 295 300

Pro Asp Leu Val Leu Asp Arg Asp Glu Gly Lys Cys Val Glu Gly Ile
305 310 315 320

25

Leu Glu Ile Phe Asp Met Leu Leu Ala Thr Thr Ser Arg Phe Arg Glu
325 330 335

30

Leu Lys Leu Gln His Lys Glu Tyr Leu Cys Val Lys Ala Met Ile Leu
340 345 350

Leu Asn Ser Ser Met Tyr Pro Leu Val Thr Ala Thr Gln Asp Ala Asp
355 360 365

35

Ser Ser Arg Lys Leu Ala His Leu Leu Asn Ala Val Thr Asp Ala Leu
370 375 380

Val Trp Val Ile Ala Lys Ser Gly Ile Ser Ser Gln Gln Gln Ser Met

385 390 395 400

Arg Leu Ala Asn Leu Leu Met Leu Leu Ser His Val Arg His Ala Arg

 405 410 415

5

Ser Ala

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CTTGGATCCA TAGCCCTGCT GTGATGAATT ACAG

34

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GATGGATCCT CACCTCAGGG CCAGGCGTCA CTG

33

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1898 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: cDNA

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CACGAATCTT TGAGAACATT ATAATGACCT TTGTGCCTCT TCTTGCAAGG TGTTTTCTCA 60
GCTGTTATCT CAAGACATGG ATATAAAAAA CTCACCATCT AGCCTTAATT CTCCTTCCTC 120
20 CTACAACTGC AGTCAATCCA TCTTACCCCT GGAGCACGGC TCCATATACA TACCTTCCTC 180
CTATGTAGAC AGCCACCATG AATATCCAGC CATGACATTC TATAGCCCTG CTGTGATGAA 240
25 TTACAGCATT CCCAGCAATG TCACTAACTT GGAAGGTGGG CCTGGTCGGC AGACCACAAG 300
CCCAAATGTG TTGTGGCCAA CACCTGGGCA CCTTTCTCCT TTAGTGGTCC ATCGCCAGTT 360
ATCACATCTG TATGCGGAAC CTCAAAAGAG TCCCTGGTGT GAAGCAAGAT CGCTAGAACA 420
30 CACCTTACCT GTAAACAGAG AGACACTGAA AAGGAAGGTT AGTGGGAACC GTTGCGCCAG 480
CCCTGTTACT GGTCCAGGTT CAAAGAGGGA TGCTCACTTC TGCGCTGTCT GCAGCGATTA 540
35 CGCATCGGGA TATCACTATG GAGTCTGGTC GTGTGAAGGA TGTAAGGCCT TTTTAAAAAG 600
AAGCATTCAA GGACATAATG ATTATATTTG TCCAGCTACA AATCAGTGTA CAATCGATAA 660
AAACCGGCGC AAGAGCTGCC AGGCCTGCCG ACTTCGGAAG TGTTACGAAG TGGGAATGGT 720

	GAAGTGTGGC TCCCGGAGAG AGAGATGTGG GTACCGCCTT GTGCGGAGAC AGAGAAGTGC	780
	CGACGAGCAG CTGCACTGTG CCGGCAAGGC CAAGAGAAGT GGC GGCCACG CGCCCCGAGT	840
5	GCGGGAGCTG CTGCTGGACG CCCTGAGCCC CGAGCAGCTA GTGCTCACCC TCCTGGAGGC	900
	TGAGCCGCCC CATGTGCTGA TCAGCCGCCC CAGTGCGCCC TTCACCGAGG CCTCCATGAT	960
10	GATGTCCCTG ACCAAGTTGG CCGACAAGGA GTTGGTACAC ATGATCAGCT GGGCCAAGAA	1020
	GATTCCCGGC TTTGTGGAGC TCAGCCTGTT CGACCAAGTG CGGCTCTTGG AGAGCTGTTG	1080
	GATGGAGGTG TTAATGATGG GGCTGATGTG GCGCTCAATT GACCACCCCG GCAAGCTCAT	1140
15	CTTTGCTCCA GATCTTGTTT TGGACAGGGA TGAGGGGAAA TCGTAGAAG GAATTCTGGA	1200
	AATCTTTGAC ATGCTCCTGG CAACTACTTC AAGGTTTCGA GAGTTAAAC TCCAACACAA	1260
20	AGAATATCTC TGTGTCAAGG CCATGATCCT GCTCAATTCC AGTATGTACC CTCTGGTCAC	1320
	AGCGACCCAG GATGCTGACA GCAGCCGGAA GCTGGCTCAC TTGCTGAACG CCGTGACCGA	1380
	TGCTTTGGTT TGGGTGATTG CCAAGAGCGG CATCTCCTCC CAGCAGCAAT CCATGCGCCT	1440
25	GGCTAACCTC CTGATGCTCC TGTCCACGT CAGGCATGCG AGTAACAAGG GCATGGAACA	1500
	TCTGCTCAAC ATGAAGTGCA AAAATGTGGT CCCAGTGTAT GACCTGCTGC TGGAGATGCT	1560
30	GAATGCCCCAC GTGCTTCGCG GGTGCAAGTC CTCCATCACG GGGTCCGAGT GCAGCCCGGC	1620
	AGAGGACAGT AAAAGCAAAG AGGGCTCCCA GAACCCACAG TCTCAGTGAC GCCTGGCCCT	1680
	GAGGTGAACT GGCCACAGA GGTCAACAAG TGAAGCGTGA ACTCCAGTGT GTCAGGAGCC	1740
35	TGGGCTTCAT CTTTCTGCTG TGTGGTCCCT CATTTGGTGA TGGCAGGCTT GGTCATGTAC	1800
	CATCCTTCCC TCCACCTTCC CAACTCTCAG GAGTCGGTGT GAGGAAGCCA TAGTTTCCCT	1860

TGTTAGCAGA GGGACATTTG AATCGAGCGT TTCCACAC

1898

(2) INFORMATION FOR SEQ ID NO: 25:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 530 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Met Asp Ile Lys Asn Ser Pro Ser Ser Leu Asn Ser Pro Ser Ser Tyr
1 5 10 15

20

Asn Cys Ser Gln Ser Ile Leu Pro Leu Glu His Gly Ser Ile Tyr Ile
 20 25 30

25

Pro Ser Ser Tyr Val Asp Ser His His Glu Tyr Pro Ala Met Thr Phe
 35 40 45

Tyr Ser Pro Ala Val Met Asn Tyr Ser Ile Pro Ser Asn Val Thr Asn
50 55 60

30

Leu Glu Gly Gly Pro Gly Arg Gln Thr Thr Ser Pro Asn Val Leu Trp
65 70 75 80

Pro Thr Pro Gly His Leu Ser Pro Leu Val Val His Arg Gln Leu Ser
 85 90 95

35

His Leu Tyr Ala Glu Pro Gln Lys Ser Pro Trp Cys Glu Ala Arg Ser
 100 105 110

Leu Glu His Thr Leu Pro Val Asn Arg Glu Thr Leu Lys Arg Lys Val

	115	120	125
	Ser Gly Asn Arg Cys Ala Ser Pro Val Thr Gly Pro Gly Ser Lys Arg		
	130	135	140
5	Asp Ala His Phe Cys Ala Val Cys Ser Asp Tyr Ala Ser Gly Tyr His		
	145	150	155 160
	Tyr Gly Val Trp Ser Cys Glu Gly Cys Lys Ala Phe Phe Lys Arg Ser		
10	165	170	175
	Ile Gln Gly His Asn Asp Tyr Ile Cys Pro Ala Thr Asn Gln Cys Thr		
	180	185	190
15	Ile Asp Lys Asn Arg Arg Lys Ser Cys Gln Ala Cys Arg Leu Arg Lys		
	195	200	205
	Cys Tyr Glu Val Gly Met Val Lys Cys Gly Ser Arg Arg Glu Arg Cys		
	210	215	220
20	Gly Tyr Arg Leu Val Arg Arg Gln Arg Ser Ala Asp Glu Gln Leu His		
	225	230	235 240
	Cys Ala Gly Lys Ala Lys Arg Ser Gly Gly His Ala Pro Arg Val Arg		
25	245	250	255
	Glu Leu Leu Leu Asp Ala Leu Ser Pro Glu Gln Leu Val Leu Thr Leu		
	260	265	270
30	Leu Glu Ala Glu Pro Pro His Val Leu Ile Ser Arg Pro Ser Ala Pro		
	275	280	285
	Phe Thr Glu Ala Ser Met Met Met Ser Leu Thr Lys Leu Ala Asp Lys		
	290	295	300
35	Glu Leu Val His Met Ile Ser Trp Ala Lys Lys Ile Pro Gly Phe Val		
	305	310	315 320
	Glu Leu Ser Leu Phe Asp Gln Val Arg Leu Leu Glu Ser Cys Trp Met		

325

330

335

Glu Val Leu Met Met Gly Leu Met Trp Arg Ser Ile Asp His Pro Gly
340 345 350

5

Lys Leu Ile Phe Ala Pro Asp Leu Val Leu Asp Arg Asp Glu Gly Lys
355 360 365

10

Cys Val Glu Gly Ile Leu Glu Ile Phe Asp Met Leu Leu Ala Thr Thr
370 375 380

Ser Arg Phe Arg Glu Leu Lys Leu Gln His Lys Glu Tyr Leu Cys Val
385 390 395 400

15

Lys Ala Met Ile Leu Leu Asn Ser Ser Met Tyr Pro Leu Val Thr Ala
405 410 415

Thr Gln Asp Ala Asp Ser Ser Arg Lys Leu Ala His Leu Leu Asn Ala
420 425 430

20

Val Thr Asp Ala Leu Val Trp Val Ile Ala Lys Ser Gly Ile Ser Ser
435 440 445

25

Gln Gln Gln Ser Met Arg Leu Ala Asn Leu Leu Met Leu Leu Ser His
450 455 460

Val Arg His Ala Ser Asn Lys Gly Met Glu His Leu Leu Asn Met Lys
465 470 475 480

30

Cys Lys Asn Val Val Pro Val Tyr Asp Leu Leu Leu Glu Met Leu Asn
485 490 495

Ala His Val Leu Arg Gly Cys Lys Ser Ser Ile Thr Gly Ser Glu Cys
500 505 510

35

Ser Pro Ala Glu Asp Ser Lys Ser Lys Glu Gly Ser Gln Asn Pro Gln
515 520 525

Ser Gln

530

(2) INFORMATION FOR SEQ ID NO: 26:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10

- (ii) MOLECULE TYPE: other nucleic acid

15

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GTGCGGATCC TCTCAAGACA TGGATATAAA

30

20

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25

- (ii) MOLECULE TYPE: other nucleic acid

30

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

35

AGTAACAGGG CTGGCGCAAC GGTTC

25

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: other nucleic acid

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ACTGGCGATG GACCACTAAA GG

22

15